

EVIDENCE BASED CHILD HEALTH 3

Understanding systematic reviews and meta-analysis

A K Akobeng

Arch Dis Child 2005;90:845–848. doi: 10.1136/adc.2004.058230

This review covers the basic principles of systematic reviews and meta-analyses. The problems associated with traditional narrative reviews are discussed, as is the role of systematic reviews in limiting bias associated with the assembly, critical appraisal, and synthesis of studies addressing specific clinical questions. Important issues that need to be considered when appraising a systematic review or meta-analysis are outlined, and some of the terms used in the reporting of systematic reviews and meta-analyses—such as odds ratio, relative risk, confidence interval, and the forest plot—are introduced.

how they selected, assessed, and analysed the primary studies, thereby not allowing readers to assess potential bias in the review process. Narrative reviews are therefore often biased, and the recommendations made may be inappropriate.⁵

WHAT IS A SYSTEMATIC REVIEW?

In contrast to a narrative review, a systematic review is a form of research that provides a summary of medical reports on a specific clinical question, using explicit methods to search, critically appraise, and synthesise the world literature systematically.⁶ It is particularly useful in bringing together a number of separately conducted studies, sometimes with conflicting findings, and synthesising their results.

By providing in a clear explicit fashion a summary of all the studies addressing a specific clinical question,⁴ systematic reviews allow us to take account of the whole range of relevant findings from research on a particular topic, and not just the results of one or two studies. Other advantages of systematic reviews have been discussed by Mulrow.⁷ They can be used to establish whether scientific findings are consistent and generalisable across populations, settings, and treatment variations, or whether findings vary significantly by particular subgroups. Moreover, the explicit methods used in systematic reviews limit bias and, hopefully, will improve reliability and accuracy of conclusions. For these reasons, systematic reviews of randomised controlled trials (RCTs) are considered to be evidence of the highest level in the hierarchy of research designs evaluating effectiveness of interventions.⁸

METHODOLOGY OF A SYSTEMATIC REVIEW

The need for rigour in the preparation of a systematic review means that there should be a formal process for its conduct. Figure 1 summarises the process for conducting a systematic review of RCTs.⁹ This includes a comprehensive, exhaustive search for primary studies on a focused clinical question, selection of studies using clear and reproducible eligibility criteria, critical appraisal of primary studies for quality, and synthesis of results according to a predetermined and explicit method.^{3,9}

WHAT IS A META-ANALYSIS?

Following a systematic review, data from individual studies may be pooled quantitatively and reanalysed using established statistical methods.¹⁰ This technique is called meta-analysis. The

Health care professionals are increasingly required to base their practice on the best available evidence. In the first article of the series, I described basic strategies that could be used to search the medical literature.¹ After a literature search on a specific clinical question, many articles may be retrieved. The quality of the studies may be variable, and the individual studies might have produced conflicting results. It is therefore important that health care decisions are not based solely on one or two studies without account being taken of the whole range of research information available on that topic.

Health care professionals have always used review articles as a source of summarised evidence on a particular topic. Review articles in the medical literature have traditionally been in the form of “narrative reviews” where experts in a particular field provide what is supposed to be a “summary of evidence” in that field. Narrative reviews, although still very common in the medical field, have been criticised because of the high risk of bias, and “systematic reviews” are preferred.² Systematic reviews apply scientific strategies in ways that limit bias to the assembly, a critical appraisal, and synthesis of relevant studies that address a specific clinical question.²

THE PROBLEM WITH TRADITIONAL REVIEWS

The validity of a review article depends on its methodological quality. While traditional review articles or narrative reviews can be useful when conducted properly, there is evidence that they are usually of poor quality. Authors of narrative reviews often use informal, subjective methods to collect and interpret studies and tend to be selective in citing reports that reinforce their preconceived ideas or promote their own views on a topic.^{3,4} They are also rarely explicit about

Correspondence to:
Dr A K Akobeng,
Department of Paediatric
Gastroenterology, Central
Manchester and
Manchester Children's
University Hospitals, Booth
Hall Children's Hospital,
Charlestown Road,
Blackley, Manchester, M9
7AA, UK; tony.akobeng@
mmc.nhs.uk

Accepted 22 April 2005

Abbreviations: RCT, randomised controlled trial

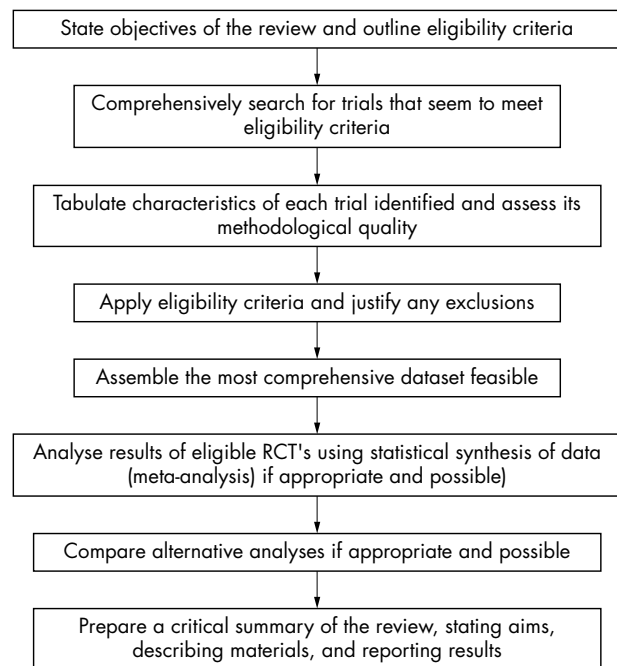


Figure 1 Methodology for a systematic review of randomised controlled trials.⁹

rationale for a meta-analysis is that, by combining the samples of the individual studies, the overall sample size is increased, thereby improving the statistical power of the analysis as well as the precision of the estimates of treatment effects.¹¹

Meta-analysis is a two stage process.¹² The first stage involves the calculation of a measure of treatment effect with its 95% confidence intervals (CI) for each individual study. The summary statistics that are usually used to measure treatment effect include odds ratios (OR), relative risks (RR), and risk differences.

In the second stage of meta-analysis, an overall treatment effect is calculated as a weighted average of the individual summary statistics. Readers should note that, in meta-analysis, data from the individual studies are not simply combined as if they were from a single study. Greater weights are given to the results from studies that provide more information, because they are likely to be closer to the “true effect” we are trying to estimate. The weights are often the inverse of the variance (the square of the standard error) of the treatment effect, which relates closely to sample size.¹² The typical graph for displaying the results of a meta-analysis is called a “forest plot”.¹³

The forest plot

The plot shows, at a glance, information from the individual studies that went into the meta-analysis, and an estimate of the overall results. It also allows a visual assessment of the amount of variation between the results of the studies (heterogeneity). Figure 2 shows a typical forest plot. This figure is adapted from a recent systematic review and meta-analysis which examined the efficacy of probiotics compared with placebo in the prevention and treatment of diarrhoea associated with the use of antibiotics.¹⁴

Description of the forest plot

In the forest plot shown in fig 2, the results of nine studies have been pooled. The names on the left of the plot are the first authors of the primary studies included. The black squares represent the odds ratios of the individual studies,

and the horizontal lines their 95% confidence intervals. The area of the black squares reflects the weight each trial contributes in the meta-analysis. The 95% confidence intervals would contain the true underlying effect in 95% of the occasions if the study was repeated again and again. The solid vertical line corresponds to no effect of treatment (OR = 1.0). If the CI includes 1, then the difference in the effect of experimental and control treatment is not significant at conventional levels ($p > 0.05$).¹⁵ The overall treatment effect (calculated as a weighted average of the individual ORs) from the meta-analysis and its CI is at the bottom and represented as a diamond. The centre of the diamond represents the combined treatment effect (0.37), and the horizontal tips represent the 95% CI (0.26 to 0.52). If the diamond shape is on the **Left** of the line of no effect, then **Less** (fewer episodes) of the outcome of interest is seen in the treatment group. If the diamond shape is on the **Right** of the line, then **more** episodes of the outcome of interest are seen in the treatment group. In fig 2, the diamond shape is found on the left of the line of no effect, meaning that less diarrhoea (fewer episodes) was seen in the probiotic group than in the placebo group. If the diamond touches the line of no effect (where the OR is 1) then there is no statistically significant difference between the groups being compared. In fig 2, the diamond shape does not touch the line of no effect (that is, the confidence interval for the odds ratio does not include 1) and this means that the difference found between the two groups was statistically significant.

APPRAISING A SYSTEMATIC REVIEW WITH OR WITHOUT META-ANALYSIS

Although systematic reviews occupy the highest position in the hierarchy of evidence for articles on effectiveness of interventions,⁸ it should not be assumed that a study is valid merely because it is stated to be a systematic review. Just as in RCTs, the main issues to consider when appraising a systematic review can be condensed into three important areas⁸:

- The validity of the trial methodology.
- The magnitude and precision of the treatment effect.
- The applicability of the results to your patient or population.

Box 1 shows a list of 10 questions that may be used to appraise a systematic review in all three areas.¹⁶

ASSESSING THE VALIDITY OF TRIAL METHODOLOGY

Focused research question

Like all research reports, the authors should clearly state the research question at the outset. The research question should include the relevant population or patient groups being studied, the intervention of interest, any comparators (where relevant), and the outcomes of interest. Keywords from the research question and their synonyms are usually used to identify studies for inclusion in the review.

Types of studies included in the review

The validity of a systematic review or meta-analysis depends heavily on the validity of the studies included. The authors should explicitly state the type of studies they have included in their review, and readers of such reports should decide whether the included studies have the appropriate study design to answer the clinical question. In a recent systematic review which determined the effects of glutamine supplementation on morbidity and weight gain in preterm babies the investigators based their review only on RCTs.¹⁷

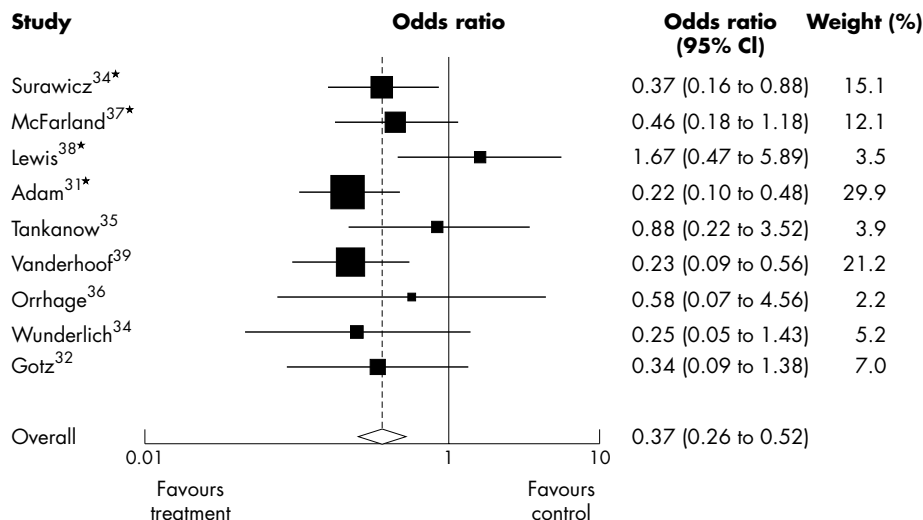


Figure 2 Effect of probiotics on the risk of antibiotic associated diarrhoea.¹⁴

Search strategy used to identify relevant articles

There is evidence that single electronic database searches lack sensitivity and relevant articles may be missed if only one database is searched. Dickersin *et al* showed that only 30–80% of all known published RCTs were identifiable using MEDLINE.¹⁸ Even if relevant records are in a database, it can be difficult to retrieve them easily. A comprehensive search is therefore important, not only for ensuring that as many studies as possible are identified but also to minimise selection bias for those that are found. Relying exclusively on one database may retrieve a set of studies that are unrepresentative of all studies that would have been identified through a comprehensive search of multiple sources. Therefore, in order to retrieve all relevant studies on a topic, several different sources should be searched to identify relevant studies (published and unpublished), and the search strategy should not be limited to the English language. The aim of an extensive search is to avoid the problem of publication bias which occurs when trials with statistically significant results are more likely to be published and cited, and are preferentially published in English language journals and those indexed in Medline.

In the systematic review referred to above, which examined the effects of glutamine supplementation on

morbidity and weight gain in preterm babies, the authors searched the Cochrane controlled trials register, Medline, and Embase,¹⁷ and they also hand searched selected journals, cross referencing where necessary from other publications.

Quality assessment of included trials

The reviewers should state a predetermined method for assessing the eligibility and quality of the studies included. At least two reviewers should independently assess the quality of the included studies to minimise the risk of selection bias. There is evidence that using at least two reviewers has an important effect on reducing the possibility that relevant reports will be discarded.¹⁹

Pooling results and heterogeneity

If the results of the individual studies were pooled in a meta-analysis, it is important to determine whether it was reasonable to do so. A clinical judgement should be made about whether it was reasonable for the studies to be combined based on whether the individual trials differed considerably in populations studied, interventions and comparisons used, or outcomes measured.

The statistical validity of combining the results of the various trials should be assessed by looking for homogeneity of the outcomes from the various trials. In other words, there should be some consistency in the results of the included trials. One way of doing this is to inspect the graphical display of results of the individual studies (forest plot, see above) looking for similarities in the direction of the results. When the results differ greatly in their direction—that is, if there is significant heterogeneity—then it may not be wise for the results to be pooled. Some articles may also report a statistical test for heterogeneity, but it should be noted that the statistical power of many meta-analyses is usually too low to allow the detection of heterogeneity based on statistical tests. If a study finds significant heterogeneity among reports, the authors should attempt to offer explanations for potential sources of the heterogeneity.

Magnitude of the treatment effect

Common measures used to report the results of meta-analyses include the odds ratio, relative risk, and mean differences. If the outcome is binary (for example, disease *v* no disease, remission *v* no remission), odds ratios or relative risks are used. If the outcome is continuous (for example, blood pressure measurement), mean differences may be used.

Box 1: Questions to consider when appraising a systematic review¹⁶

- Did the review address a clearly focused question?
- Did the review include the right type of study?
- Did the reviewers try to identify all relevant studies?
- Did the reviewers assess the quality of all the studies included?
- If the results of the study have been combined, was it reasonable to do so?
- How are the results presented and what are the main results?
- How precise are the results?
- Can the results be applied to your local population?
- Were all important outcomes considered?
- Should practice or policy change as a result of the evidence contained in this review?

ODDS RATIOS AND RELATIVE RISKS

Odds and odds ratio

The *odds* for a group is defined as the number of patients in the group who achieve the stated end point divided by the number of patients who do not. For example, the odds of acne resolution during treatment with an antibiotic in a group of 10 patients may be 6 to 4 (6 with resolution of acne divided by 4 without = 1.5); in a control group the odds may be 3 to 7 (0.43). The *odds ratio*, as the name implies, is a ratio of two odds. It is simply defined as the ratio of the odds of the treatment group to the odds of the control group. In our example, the odds ratio of treatment to control group would be 3.5 (1.5 divided by 0.43).

Risk and relative risk

Risk, as opposed to odds, is calculated as the number of patients in the group who achieve the stated end point divided by the *total* number of patients in the group. Risk ratio or relative risk is a ratio of two "risks". In the example above the risks would be 6 in 10 in the treatment group (6 divided by 10 = 0.6) and 3 in 10 in the control group (0.3), giving a risk ratio, or relative risk of 2 (0.6 divided by 0.3).

Interpretation of odds ratios and relative risk

An odds ratio or relative risk greater than 1 indicates increased likelihood of the stated outcome being achieved in the treatment group. If the odds ratio or relative risk is less than 1, there is a decreased likelihood in the treatment group. A ratio of 1 indicates no difference—that is, the outcome is just as likely to occur in the treatment group as it is in the control group.¹¹ As in all estimates of treatment effect, odds ratios or relative risks reported in meta-analysis should be accompanied by confidence intervals.

Readers should understand that the odds ratio will be close to the relative risk if the end point occurs relatively infrequently, say in less than 20%.¹⁵ If the outcome is more common, then the odds ratio will considerably overestimate the relative risk. The advantages and disadvantages of odds ratios *v* relative risks in the reporting of the results of meta-analysis have been reviewed elsewhere.¹²

Precision of the treatment effect: confidence intervals

As stated earlier, confidence intervals should accompany estimates of treatment effects. I discussed the concept of confidence intervals in the second article of the series.⁸ Ninety five per cent confidence intervals are commonly reported, but other intervals such as 90% or 99% are also sometimes used. The 95% CI of an estimate (for example, of odds ratios or relative risks) will be the range within which we are 95% certain that the true population treatment effect will lie. The width of a confidence interval indicates the precision of the estimate. The wider the interval, the less the precision. A very long interval makes us less sure about the accuracy of a study in predicting the true size of the effect. If the confidence interval for relative risk or odds ratio for an estimate includes 1, then we have been unable to demonstrate a statistically significant difference between the groups being compared; if it does not include 1, then we say that there is a statistically significant difference.

APPLICABILITY OF RESULTS TO PATIENTS

Health care professionals should always make judgements about whether the results of a particular study are applicable to their own patient or group of patients. Some of the issues that one need to consider before deciding whether to incorporate a particular piece of research evidence into clinical practice were discussed in the second article of the series.⁸ These include similarity of study population to your population, benefit *v* harm, patients preferences, availability, and costs.

CONCLUSIONS

Systematic reviews apply scientific strategies to provide in an explicit fashion a summary of all studies addressing a specific question, thereby allowing an account to be taken of the whole range of relevant findings on a particular topic. Meta-analysis, which may accompany a systematic review, can increase power and precision of estimates of treatment effects. People working in the field of paediatrics and child health should understand the fundamental principles of systematic reviews and meta-analyses, including the ability to apply critical appraisal not only to the methodologies of review articles, but also to the applicability of the results to their own patients.

Competing interests: none declared

REFERENCES

- 1 Akobeng AK. Evidence based child health 1. Principles of evidence based medicine. *Arch Dis Child* 2005;**90**:837–40.
- 2 Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med* 1997;**126**:376–80.
- 3 Pai M, McCulloch M, Gorman JD, et al. Systematic reviews and meta-analyses: an illustrated, step-by-step guide. *Natl Med J India* 2004;**17**:86–95.
- 4 McGovern DPB. Systematic reviews. In: McGovern DPB, Valori RM, Summerskill WSM, eds. *Key topics in evidence based medicine*. Oxford: BIOS Scientific Publishers, 2001:17–9.
- 5 McAlister FA, Clark HD, van Walraven C, et al. The medical review article revisited: has the science improved? *Ann Intern Med* 1999;**131**:947–51.
- 6 Sackett DL, Strauss SE, Richardson WS, et al. *Evidence-based medicine: how to practice and teach EBM*. London: Churchill Livingstone, 2000.
- 7 Mulrow CD. Systematic reviews: rationale for systematic reviews. *BMJ* 1994;**309**:597–9.
- 8 Akobeng AK. Evidence based child health 2. Understanding randomised controlled trials. *Arch Dis Child* 2005;**90**:840–4.
- 9 Greenhalgh T. How to read a paper: papers that summarise other papers (systematic reviews and meta-analyses). *BMJ* 1997;**315**:672–5.
- 10 Muir Gray JA. *Evidence based healthcare. How to make health policy and management decisions*. London: Churchill Livingstone, 2001:125–6.
- 11 Lang TA, Secic M. *How to report statistics in medicine*. Philadelphia: American College of Physicians, 1997.
- 12 Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Smith GD, Altman DG, eds. *Systematic reviews in healthcare: meta-analysis in context*. London: BMJ Publishing Group, 2001:285–312.
- 13 Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. *BMJ* 2001;**322**:1479–80.
- 14 D'Souza AL, Rajkumar C, Cooke J, et al. Probiotics in prevention of antibiotic associated diarrhoea: meta-analysis. *BMJ* 2002;**324**:1361.
- 15 Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. *BMJ* 1997;**315**:1533–7.
- 16 **Critical Appraisal Skills Programme**. Appraisal Tools. Oxford, UK. <http://www.phru.nhs.uk/casp/appraisa.htm> (accessed 10 Dec 2004).
- 17 Tubman TRJ, Thompson SW. Glutamine supplementation for prevention of morbidity in preterm infants. The Cochrane Database of Systematic Reviews 2001, Issue 4.
- 18 Dickersin K, Scherer R, Lefebvre C. Systematic reviews: identifying relevant studies for systematic reviews. *BMJ* 1994;**309**:1286–91.
- 19 Clarke M, Oxman AD, eds. Selecting studies. Cochrane reviewers' handbook 4.2.0 [updated March 2003]. In: *The Cochrane library*, issue 2. Oxford: Update Software, 2003.